

Repositório Inf. Bayseana

Daniel Figueira Ferreira

2023-09-01

Table of contents

Preface	3
I Entregas relacionadas as aulas práticas	4
1 Relatório: Acompanhamento pós COVID-19	6
2 Indicação de acompanhamento de pacientes pós COVID-19	7
2.1 Análise descritiva	7
2.2 TC.Inicial	7
2.3 TC.Final	7
2.4 Matriz De Confusão	7
2.5 Quadro do paciente	8
2.6 Análise: Quadro em relação ao Tabagismo	8
2.6.1 Hipótese: Tabagismo é um fator de risco ?	11
2.7 Análise: Quadro em relação a Gravidade	12
2.7.1 Hipótese: A Gravidade é um fator de risco ?	14
3 Relatório 2: Número de mortes por coice de cavalo	15
3.1 Introdução	15
3.2 Objetivos	15
3.3 Desenvolvimento	16
4 Relatorio 3: Ácaro-vermelho-europeu	20
4.1 Introdução	20
4.2 Laboratório	20
4.2.1 Modelo Poisson	21
4.2.2 Modelo Binomial Negativa	21
5 Relatorio 4: Tornados	24
References	27

Preface

This is a Quarto book.

To learn more about Quarto books visit <https://quarto.org/docs/books>.

1 + 1

[1] 2

Part I

Entregas relacionadas as aulas práticas

Cada página está relacionada com um relatório entregue.

1 Relatório: Acompanhamento pós COVID-19

2 Indicação de acompanhamento de pacientes pós COVID-19

Um projeto submetido ao LabEst teve o objetivo de verificar se há indicação de acompanhamento de pacientes pós Covid-19 em relação a gravidade da tomografia computadorizada de tórax relacionada à função pulmonar, devido à possibilidade da persistência de alterações sequelares pulmonares a partir dos exames.

2.1 Análise descritiva

Temos 4 colunas nesses dados, **Tabagismo**, **Gravidade**, **TC.inicial** e **TC.final**.

- Tabagismo: Variável dicotômica indicando se há histórico de tabagismo.
- Gravidade: Variável dicotômica indicando se houveram complicações com o paciente durante a doença, como intubação.
- TC.inicial: variável categórica com a indicação da gravidade diagnosticada via exame de imagem, realizada enquanto o paciente estava com COVID-19.
- TC.final: Variável categórica com a indicação da gravidade diagnosticada via exame de imagem, realizada com pelo menos 60 dias após o término da doença COVID-19.

Vamos analisar os dados para TC.inicial e TC.final.

2.2 TC.Inicial

2.3 TC.Final

2.4 Matriz De Confusão

Table 2.1: Matriz de Confusão

	Nenhum	Minimo	Leve	Moderado	Sequela
Nenhum	6	5	0	0	0

	Nenhum	Minimo	Leve	Moderado	Sequela
Minimo	23	29	0	0	0
Leve	11	65	9	0	0
Moderado	0	25	23	1	4
Grave	0	3	0	0	0

Como foi possível ver, temos poucos dados para cada uma das gravidades, nas variáveis categóricas TC.inicial e TC.final, logo foi preciso utilizar um método alternativo para olhar o problema, pois caso contrário não haveria amostras o suficiente para as análises, ficariam imprecisas. A maneira encontrada foi abordar o problema de modo a ter apenas 3 classes para o quadro pulmonar do indivíduo, igual ou estável, melhora e piora. Aqueles que estão na diagonal principal serão vistos como os que mantêm sua condição como Igual ou estável, os que estão abaixo da diagonal principal representam os indivíduos que melhoraram e por fim aqueles acima da diagonal principal representam os que tiveram piora de quadro.

2.5 Quadro do paciente

2.5.0.0.1 Estimativas pontuais

O valor médio estimado para cada classe foi de:

Table 2.2: Média para o quadro pulmonar do indivíduo.

Igual	Melhora	Piora
0.2222224	0.7294691	0.04830851

2.6 Análise: Quadro em relação ao Tabagismo

Agora é preciso investigar se dado o histórico de tabagismo da pessoa, o indivíduo será afetado ou não.

2.6.0.1 Não fumantes

Table 2.3: Tabela de contigência entre TC.inicial e TC.final

	Nenhum	Minimo	Leve	Moderado	Sequela
Nenhum	6	3	0	0	0

	Nenhum	Minimo	Leve	Moderado	Sequela
Minimo	18	21	0	0	0
Leve	7	41	6	0	0
Moderado	0	20	17	0	1
Grave	0	3	0	0	0

2.6.0.2 Fumantes

Table 2.4: Tabela de contingência entre TC.inicial e TC.final

	Nenhum	Minimo	Leve	Moderado	Sequela
Nenhum	0	2	0	0	0
Minimo	5	8	0	0	0
Leve	4	24	3	0	0
Moderado	0	5	6	1	3
Grave	0	0	0	0	0

[1] 0.943986

2.6.0.3 Estimativas

Table 2.5: Resultado para o quadro dos pacientes, sem histórico de tabagismo em porcentagem

	Estimativa	Desvio padrão
Estável	23.29	3.49
Melhora	73.29	3.65
Piora	3.42	6.98

Table 2.6: Resultados para o quadro dos pacientes, dado o histórico positivo para tabagismo em porcentagem

	Estimativa	Desvio padrão
Estável	20.32	4.99
Melhora	70.31	5.67
Piora	9.38	9.98

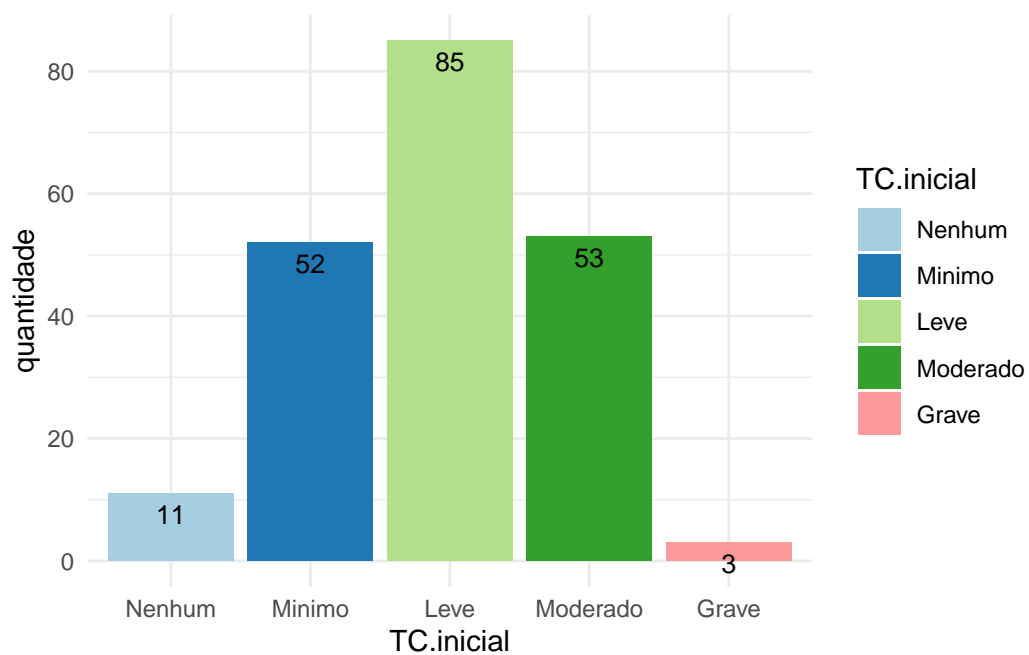


Figure 2.1: Grafico de barras para a variável categórica TC.inicial

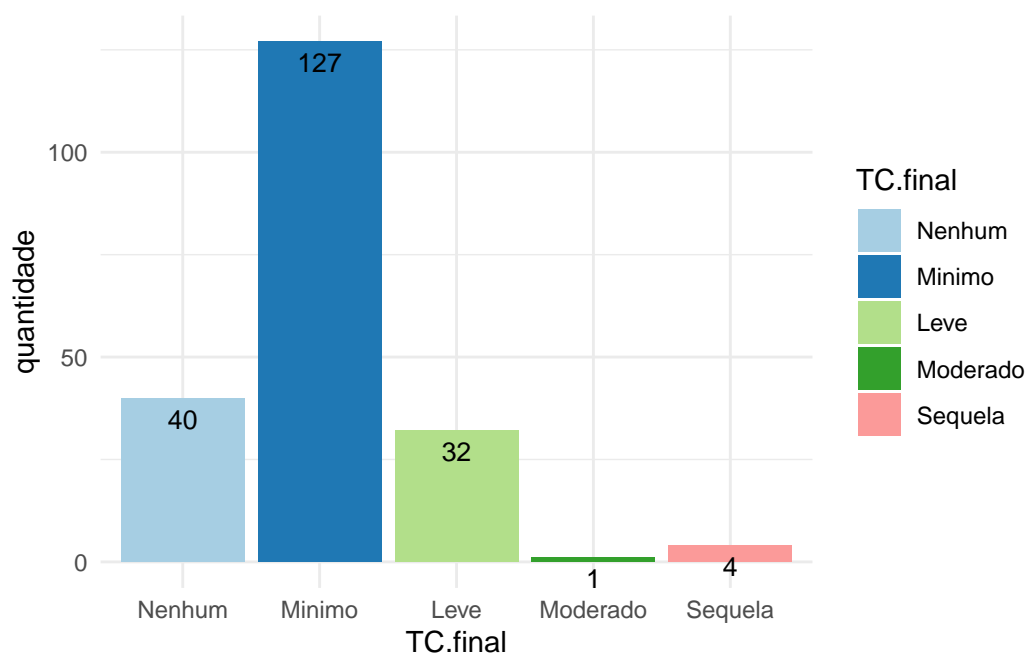


Figure 2.2: Gráfico de barras para a variável categórica TC.final

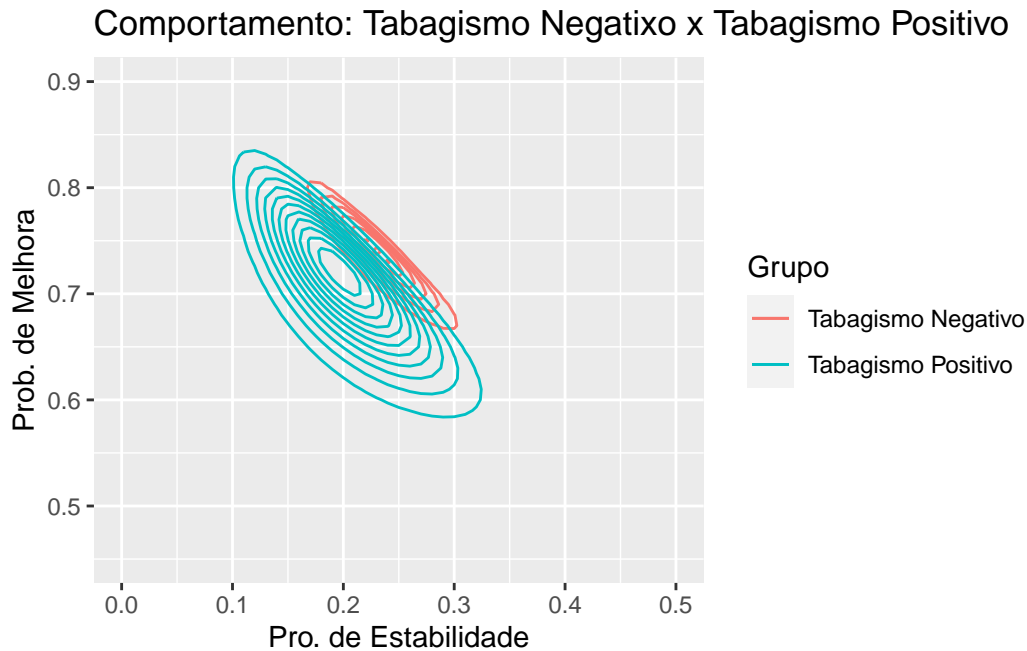


Figure 2.3: Comportamentos de cada grupo por sua devida probabilidade

2.6.0.4 Gráficos

Primeiramente chamamos a atenção, para a seção Estimativas, onde pode-se ver que pessoas sem histórico de tabagismo possuem uma menor tendência de piora no seu quadro pulmonar, enquanto que aqueles que possuem histórico, possuem 3x mais chance de piorarem de quadro. Importante notar também que tanto para melhora quanto para estabilidade, não se tem uma grande diferença, algo que está evidente na seção gráficos.

Utilizando um gráfico, é possível ver que os dados estão bem agrupados, com a diferença que aqueles que possuem histórico de tabagismo possuem uma maior variância. Sendo assim, não é aconselhável o acompanhamento para esses grupos, pois as pessoas tem uma tendência de melhorar, e para ambos os grupos essa tendência é relativamente homogênea.

2.6.1 Hipótese: Tabagismo é um fator de risco ?

Baseado nos dados apresentados, as estimativas dão indícios de que pessoas com histórico de tabagismo possuem um certo grau de piora no seu quadro pulmonar, sendo assim é necessários descobrir, se dado o histórico positivo qual a probabilidade de um indivíduo qualquer, possuir uma piora no seu quadro pulmonar.

Hipótese : Possui Historico < Não Possui Historico

Baseado na hipótese acima, o esperado é que os pacientes com o histórico positivo de tabagismo, tenham uma queda pequena na taxa para o quadro estável e o de melhora, ou seja ocorre um crescimento na taxa de piora. A taxa de piora foi apurada como 94% para indivíduos com histórico positivo para tabagismo.

Sendo assim, temos evidências que nos levam a acreditar que o tabagismo é um fator de risco a ser observado, durante a infecção da COVID-19.

2.7 Análise: Quadro em relação a Gravidade

Agora checaremos se deve ser aconselhado acompanhamento a pacientes que possuíram complicações durante o período de infecção da COVID-19.

2.7.0.1 Sem Gravidade

Table 2.7: Tabela de confusão para pacientes com Gravidade negativa

	Nenhum	Minimo	Leve	Moderado	Sequela
Nenhum	5	2	0	0	0
Minimo	21	22	0	0	0
Leve	6	28	5	0	0
Moderado	0	3	5	0	0
Grave	0	0	0	0	0

2.7.0.2 Com Gravidade

Table 2.8: Tabela de contingência para pacientes com Gravidade positiva

	Nenhum	Minimo	Leve	Moderado	Sequela
Nenhum	1	3	0	0	0
Minimo	2	7	0	0	0
Leve	5	37	4	0	0
Moderado	0	22	18	1	4
Grave	0	3	0	0	0

2.7.0.3 Estimativas

Table 2.9: Resultados para o quadro dos pacientes, que tiveram Gravidade como negativa

	Estimativa	Desvio padrão
Estável	33.00	4.68
Melhora	63.99	4.78
Piora	3.01	9.36

Table 2.10: Resultados para o quadro dos pacientes, que tiveram Gravidade como positiva

	Estimativa	Desvio padrão
Estável	12.73	3.16
Melhora	80.00	3.80
Piora	7.27	6.32

2.7.0.4 Gráfico

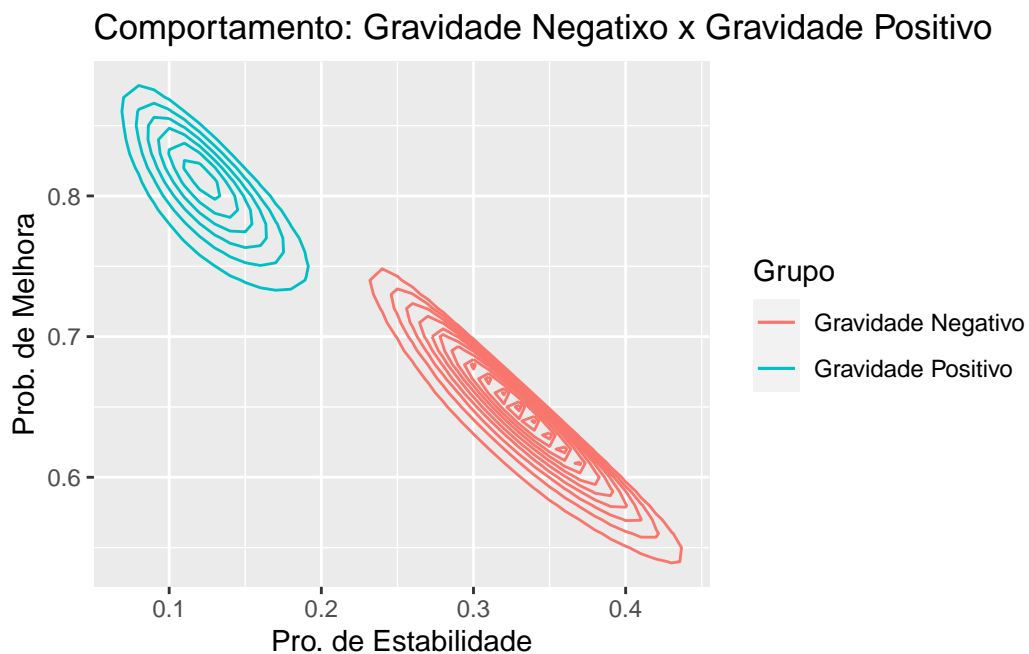


Figure 2.4: Comportamento para cada grupo de acordo com suas probabilidades

Aqui é importante ressaltar que para os grupos de gravidade, cada exame registrado na variável TC.inicial e TC.final, possuem uma diferença de 60 dias, nesse período quem tem uma complicação possui uma maior chance de sair dela, ou seja, deixa a condição estável e passa para a de melhora, poucos serão aqueles que não irão melhorar.

Por essa razão é possível notar uma relação inversa entre a condição estável e a condição melhora, adicionado ao fato que pessoas que estão com Gravidade negativa, possuem uma tendência de continuarem estáveis, por isso uma menor variância é registrada para esse grupo. Sendo assim nos resta chamar atenção a uma taxa de piora de 7% para os que possuem Gravidade positiva, e por fim não aconselhamos o acompanhamento desses pacientes.

2.7.1 Hipótese: A Gravidade é um fator de risco ?

Estimando a partir dos dados, foi possível notar na tabela 9 que possuímos um grau maior de piora nos pacientes entubados. Agora é preciso saber qual a probabilidade de um paciente piorar devido a ocorrência de complicações durante o tratamento da COVID-19, ou seja:

Hiptese : Com Complicaes < Sem Complicaes

Essa hipótese se baseia no fato de que, caso o paciente não tenha melhorado, ou se mantido estável, então o indivíduo só pode ter piorado o seu quadro. Logo é possível afirmar que aqueles com registro de complicação possuem 91% de probabilidade de terem uma piora no seu quadro pulmonar, sendo assim temos evidências de que Gravidade é um fator de risco.

3 Relatório 2: Número de mortes por coice de cavalo

3.1 Introdução

Em seu famoso livro *Lei dos Pequenos Números*, Bortkiewicz (1898) mostra como o número de mortes por coice de cavalo se aproxima do modelo Poisson. Os dados originais foram organizados em uma tabela de dupla entrada, com os anos nas linhas (de 1875 a 1894, compreendendo 20 anos) e os batalhões nas colunas (16 ao todo). Os dados que iremos analisar (`HorseKicks`, do pacote `vcd`) foram compilados por Fisher (1925), considerando dez batalhões com o mesmo tipo de organização. Cada cela da tabela (batalhão-ano) é considerada uma realização de uma variável aleatória, o que implica em uma amostra de tamanho 200.

Table 3.1: Frequência de morte por coice de cavalo

n_deaths	0	1	2	3	4
frequencia	109	65	22	3	1

3.2 Objetivos

Para este relatório os seguintes pontos devem ser realizados.

1. Mostrar que a priori $Gama(a, b)$ é conjugada para o modelo $Poisson(\lambda)$.
2. Escolher valores razoáveis para a e b de modo a ter uma priori vaga.
3. Encontrar uma estimativa pontual e intervalar para λ
4. Realizar uma análise comparando as frequências relativas simuladas a partir da preditiva a posteriori com as frequências da amostra original.
5. Encontrar a priori de Jeffreys, mostrar que a posteriori é própria e repetir os passos anteriores.

3.3 Desenvolvimento

1. Mostrar que a priori $Gama(a, b)$ é conjugada para o modelo $Poisson(\lambda)$.

Para isso, provamos que a poisson pertence a família exponencial, onde assume a forma $L(\lambda) = \frac{1}{\prod x!} e^{-\lambda n} e^{\sum x \log \lambda}$. Onde obtemos os membros que compõem a família exponencial:

$$\begin{aligned} h(x) &= \frac{1}{\prod x} \\ a(\lambda) &= e^{\lambda n} \\ w(\lambda) &= \log \lambda \\ T(x) &= \sum x \end{aligned} \tag{3.1}$$

Sendo assim provamos que a poisson pertence a família exponencial.

Deixando a verossimilhança apenas em função de λ ficamos com $L(\lambda) \propto \lambda^{\sum x_i} e^{-\lambda n}$, essa que por sua vez possui um núcleo semelhante ao da $\lambda \sim Gama(a, b)$ que é dado abaixo por:

$$f(\lambda) \propto \lambda^{a-1} e^{-b\lambda}$$

Logo provamos que o modelo conjugado da Poisson é a Gama, e que resulta na posteriori:

$$\lambda|x \sim Gama(a + \sum x_i, b + n)$$

O $\sum x$ corresponde ao numero de pessoas que vieram a falecer, e o somatório da multiplicação entre a quantidade de ocorrências pelo seu nDeaths é o resultado para a estatística suficiente.

2. Escolher valores razoáveis para a e b de modo a ter uma priori vaga.

Os valores escolhidos foram $a = 0.001$ e $b = 0.001$.

$$\lambda \sim Gama(a, b)$$

Média e variância para a priori:

$$\begin{aligned} E(\lambda) &= \frac{a}{b} = \frac{0.001}{0.001} = 1 \\ Var(\lambda) &= \frac{a}{b^2} = \frac{0.001}{0.001^2} = 1000 \end{aligned} \tag{3.2}$$

3. Encontrar uma estimativa pontual e intervalar para λ a posteriori.

A estimativa pontual para λ é:

$$\begin{aligned} E(\lambda|x) &= \frac{\alpha}{\beta} = \frac{a + \sum x_i}{b + n} = \frac{122.001}{200.001} = 0.6100019 \\ Var(\lambda|x) &= \frac{\alpha}{\beta^2} = \frac{a + \sum x_i}{(b + n)^2} = \frac{122.001}{200.001^2} = 0.003049994 \end{aligned} \quad (3.3)$$

Intervalo de Credibilidade:

Table 3.2: Intervalo de Credibilidade para lambda

2.5%	97.5%
0.5065702	0.7228996

4. Realizar uma análise comparando as frequências relativas simuladas a partir da preditiva a posteriori com as frequências da amostra original.

Pelo gráfico é possível ver que as frequências relativas registradas pela simulação estão de acordo com as frequências obtidas da amostra original, sendo assim podemos dizer que a poisson é um modelo adequado.

5. Encontrar a priori de Jeffreys, mostrar que a posteriori é própria e repetir os passos anteriores. A posteriori encontrada foi

A priori de Jeffreys obtida para a poisson foi:

$$f(\lambda) = \lambda^{-1/2}$$

Para a posteriori ficamos com:

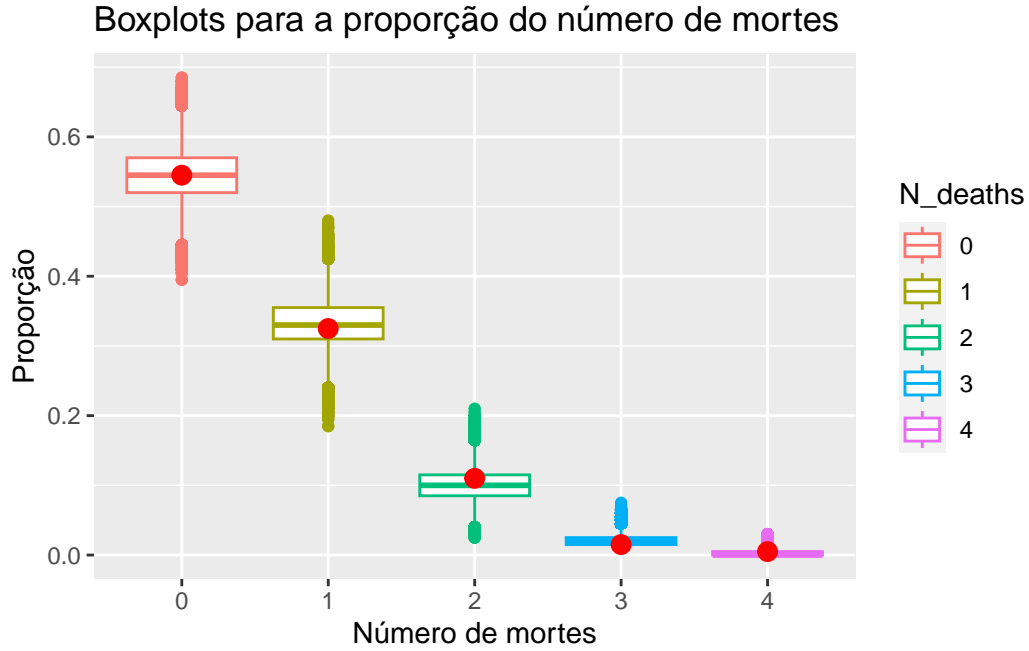


Figure 3.1: Boxplots para a proporção do número de mortes por coice de cavalo

$$\begin{aligned}
 f(\lambda) &= \frac{L(\lambda)f(\lambda)}{f(x)} \\
 &\propto L(\lambda)f(\lambda) \\
 &= \frac{e^{-\lambda n} \lambda^{\sum x_i}}{\prod x_i} \lambda^{-\frac{1}{2}} \\
 &\propto \lambda^{\sum x_i + \frac{1}{2} - 1} e^{-\lambda n}
 \end{aligned} \tag{3.4}$$

Logo, a partir de uma priori imprópria, obtivemos uma posteriori própria com distribuição $Gama(\sum x_i + 0.5, n)$.

Estimativa pontual para λ :

$$\begin{aligned}
 E(\lambda|x) &= \frac{\alpha}{\beta} = \frac{0.5 + \sum x_i}{n} = \frac{122.5}{200} = 0.6125 \\
 Var(\lambda|x) &= \frac{\alpha}{\beta^2} = \frac{0.5 + \sum x_i}{n^2} = \frac{122.5}{200^2} = 0.0030625
 \end{aligned} \tag{3.5}$$

Intervalo de Credibilidade:

Table 3.3: Intervalo de credibilidade para lambda com priori de Jeffrey

2.5%	97.5%
0.5088464	0.7256196

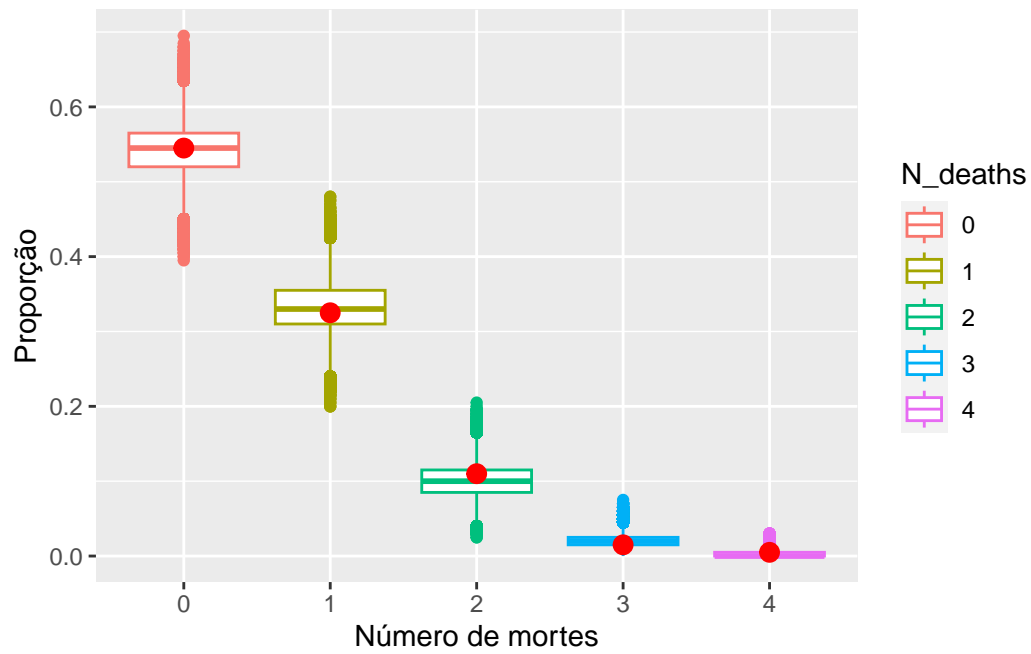


Figure 3.2: Boxplots para a proporção do número de mortes por coice de cavalo

Novamente é possível ver que o modelo poisson é adequado para os dados.

4 Relatório 3: Ácaro-vermelho-europeu

4.1 Introdução

Em 18 de julho de 1951, 25 folhas foram selecionadas aleatoriamente de cada uma das seis árvores McIntosh em um único pomar que recebeu o mesmo tratamento de pulverização, e o número de fêmeas adultas foi contado em cada folha e pode ser visto na tabela abaixo.

4.2 Laboratório

Construa uma análise baseada na distribuição Poisson. Realize uma análise da preditiva a posteriori e Verifique que o modelo Poisson não é adequado para este problema. Em seguida:

1. Seja $X|\rho, \phi$ o número de de ácaros/folha, onde ϕ é conhecido, mostre que $\rho \sim \text{Beta}(a, b)$, cuja função de probabilidade é dada por

$$P(x|\rho, \phi) = \frac{\Gamma(\phi + x)}{x!\Gamma(\phi)} \rho^\phi (1 - \phi)^x \quad (4.1)$$

onde $\rho \in (0, 1)$, $\phi > 0$ e $x \in \mathbb{N}$. Supondo que ϕ é conhecido, mostre que $\rho \sim \text{Beta}(a, b)$ é uma priori conjugada para esse modelo, concluindo que:

$$\rho|\phi, x \sim \text{Beta}(\phi n + a, \sum_{i=1}^n x_i + b) \quad (4.2)$$

2. Seja $\pi(\phi)$ a priori para ϕ . Prove que

Table 4.1: Tabela de frequência

No. acaros/folha	0	1	2	3	4	5	6	7	8
Frequencia	70	38	17	10	9	3	2	1	0

$$f(\phi) \propto \pi(\phi) B(\phi n + a, \sum_{i=1}^n x_i + b) \Gamma(\phi)^{-1} \prod_{i=1}^n \Gamma(\phi + x_i) \quad (4.3)$$

3. Encontre uma priori razoável para ϕ . Faça a análise preditiva a posteriori e mostre que o modelo é adequado. Dê estimativas pontuais e intervalares para ρ e ϕ .

4.2.1 Modelo Poisson

Simulando a preditiva a posteriori a partir de um modelo Poisson, é possível ver que o modelo não se adequa aos dados, sendo os pontos a frequência observada e os boxplots as respectivas frequências simuladas. É possível observar que na observação 0, onde nenhum ácaro foi encontrado, a frequência simulada está muito distante da observada, e se no ponto de maior densidade, temos essa discrepância. Então o modelo poisson não é o correto para explicar os dados.

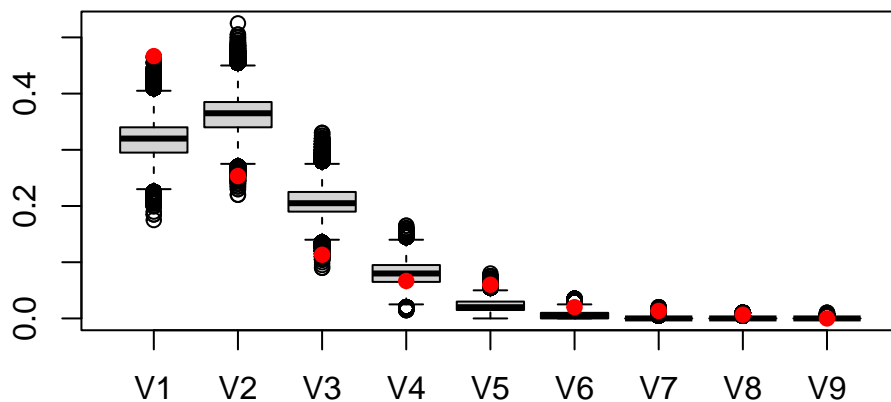


Figure 4.1: Preditiva a posteriori para a Poisson

4.2.2 Modelo Binomial Negativa

Primeiramente devemos provar que $\rho \sim \text{Binomial Negativa}(a, b)$. Onde $\Theta = \rho, \phi$

$$P(x|\Theta) \propto L(\Theta)P(\Theta) \quad (4.4)$$

$$L(\Theta) = \frac{\prod \Gamma(\phi + x_i)}{\prod x_i! \Gamma(\phi)^n} \rho^{n\phi} (1 - \rho)^{\sum x_i} \quad (4.5)$$

Para ρ faremos a priori sendo $Beta(a, b)$ e utilizaremos ϕ como constante.

$$\begin{aligned} P(\rho|x, \phi) &\propto \frac{\prod \Gamma(\phi + x_i)}{\prod x_i! \Gamma(\phi)^n} \rho^{n\phi} (1 - \rho)^{\sum x_i} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \rho^{a-1} (1 - \rho)^{b-1} \\ &\propto \rho^{n\phi+a-1} (1 - \rho)^{\sum x_i+b-1} \end{aligned} \quad (4.6)$$

Encontramos o núcleo de uma distribuição Beta, assim provando que possui $\rho|\phi, x \sim Beta(n\phi + a, \sum x_i + b)$.

Olhando a posteriori de ρ , é fácil ver que essa depende de ϕ , e para se obter ϕ é necessário integrar em ρ . Assim obteremos uma posteriori que não depende de ρ .

Para $\pi(\phi)$ como sendo a priori de ϕ devemos integrar em ρ , e assim obteremos a marginal de ϕ .

$$\begin{aligned} P(\Theta|x) &\propto \frac{\prod \Gamma(\phi + x_i)}{\Gamma(\phi)^n} \rho^{n\phi+a-1} (1 - \rho)^{\sum x_i+b-1} \frac{1}{B(n\phi + a, \sum x_i + b)} \pi(\phi) \\ P(\phi|x) &\propto \frac{\prod \Gamma(\phi + x_i)}{\Gamma(\phi)^n} \pi(\phi) Beta(n\phi + a, \sum x_i + b) \end{aligned} \quad (4.7)$$

Sendo assim, concluímos o ponto 2. Agora para a priori de ϕ faremos $\pi(\phi) \sim Gama(c, d)$ com o objetivo de obter a preditiva a posteriori.

A preditiva posteriori para o modelo da Binomial Negativa na Figure 4.2 possui a menor distância entre as frequências observadas e a frequência simulada do modelo, e ao compararmos com a Figure 4.1 que foi considerada inadequada, temos a comprovação de que o modelo da Binomial Negativa é aquele que melhor explica os dados provenientes da Table 4.1.

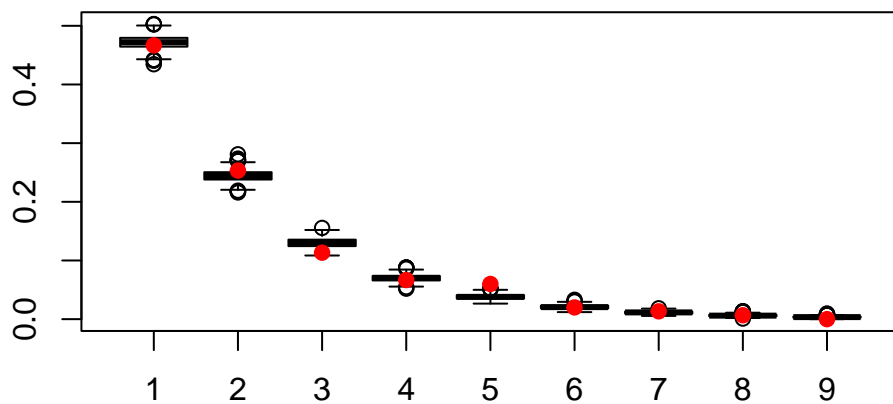


Figure 4.2: Preditiva a posteriori para a Binomial Negativa

Table 4.2: Estimação Pontual para os parâmetros

	rho	phi
Media	0.4540824	0.9542766
Erro	0.0343709	0.0752230

Table 4.3: Intervalo de credibilidade para rho e phi

	2.5%	97.5%
rho	0.3869666	0.5204929
phi	0.8111138	1.1001647

5 Relatorio 4: Tornados

Abaixo, segue o número anual de tornados em Lafayette Parish, Louisiana, entre 1950 e 2012.

Table 5.1: Tabela de frequência absoluta para os tornados

0	1	2	3
35	19	6	3

Para esses dados, temos interesse em explicar a frequência de tornados registrados durante o ano ou seja trata-se de um modelo de contagem, o modelo mais básico para modelar dados de contagem é o modelo Poisson.

Para a aplicação do modelo Poisson, faremos uso de uma Verossimilhança Poisson, com uma priori Gama como conjugada, abaixo é possível ver a verossimilhança e priori para modelo Poisson:

$$\begin{aligned} L(\theta) &= \frac{e^{-n\theta} \theta^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i} \\ \pi(\theta) &= \text{Gama}(r, \frac{p}{1-p}) \end{aligned} \tag{5.1}$$

Ao marginalizar em θ obtemos uma Binomial negativa onde $Y \sim NB(\sum x, \frac{n}{1+n})$

Obtendo a densidade, temos a proporção para o número de tornados durante ao ano, que pode ser vizualizado abaixo em Figure 5.1.

Como é possível ver, a princípio poderíamos vir a dizer que o modelo ajusta bem os dados, mas existe uma atenção que deve ser vista, o modelo criado subestima ou superestima a frequênciade de tornados em todos os 4 pontos.

Isso ocorre devido a uma maior presença de zeros na base de dados, em um modelo poisson tradicional, apesar de existir probabilidade no ponto zero quando λ é pequeno, é possível que os dados não estejam sendo provenientes de uma distribuição Poisson, mas de que sejam modelados por uma mistura onde o zero é o contaminante, o zero está sendo inflacionado.

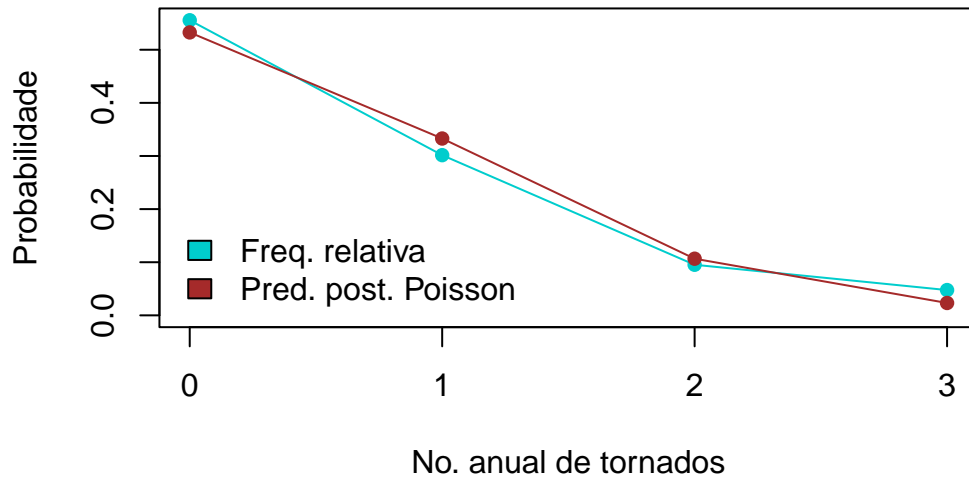


Figure 5.1: Comparação entre a frequência observada e a predita pelo modelo Poisson

Logo poderíamos modelar com um modelo Poisson inflada de zeros, a inflada por zeros ocorre quando uma parcela dos dados possui distribuição $Y|\lambda \sim Poisson(\lambda)$ e outra possui $Z|p \sim Bernoulli(p)$, quando $z = 0|x_i = 0$ teremos o valor sendo proveniente da Poisson, caso $z = 1|x_i = 0$ teremos a ocorrência de zero.

Ajustando o modelo por meio do amostrador de Gibbs, temos os seguintes resultados na Figure 5.2. A preditiva posteriori da Poisson inflacionada por zeros consegue explicar a ocorrência anual de tornados em Lafayette Parish, Louisiana no período de 1950 a 2012.

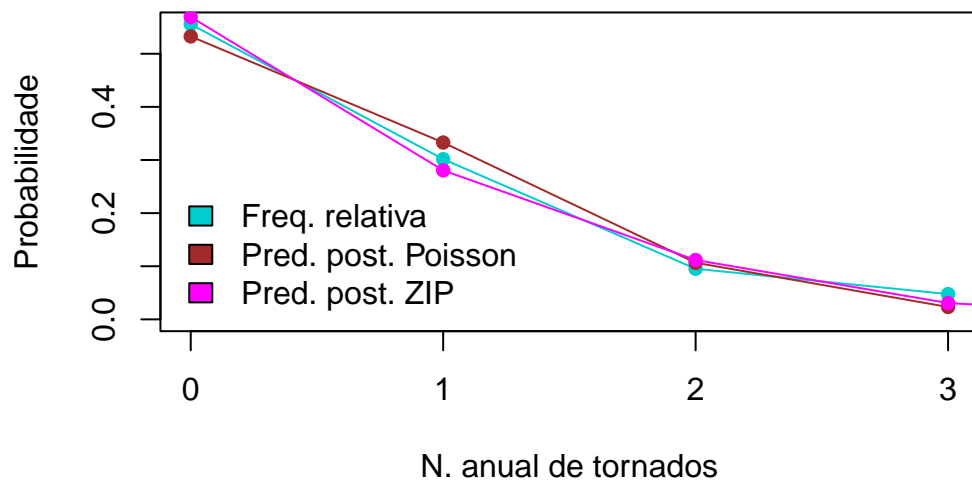


Figure 5.2: Comparação entre a frequência observada e as simuladas do modelo Poisson e ZIP

References